A Physical Charge-Controlled Model for MOS Transistors

Mary Ann Maher and Carver A. Mead
California Institute of Technology
Pasadena, California 91125

Abstract

As MOS devices scale to submicron lengths, short-channel effects become more pronounced, and an improved transistor model becomes a necessary tool for the VLSI designer [10]. We present a simple, physically based charge-controlled model. The current in the MOS transistor is described in terms of the mobile charge in the channel, and incorporates the physical processes of drift and diffusion. The effect of velocity saturation is included in the drift term. We define a complete set of natural units for velocity, voltage, length, charge, and current. The solution of the dimensionless current-flow equations using these units is a simple continuous expression, equally applicable in the subthreshold, saturation, and "ohmic" regions of transistor operation, and suitable for computer simulation of integrated circuits. The model is in agreement with measurements on short-channel transistors down to $0.35\,\mu$ channel length.

General Approach

We will begin by obtaining the channel current for a transistor in saturation. This condition is equivalent to the assumption that the mobile charge at the drain is moving at the saturated velocity v_0. Our strategy will be to choose a value for the mobile charge per unit area at the barrier maximum near the source. This value Q_s is obtained by integrating the Fermi distribution in the source times the density of states in the channel region with respect to energy. Given the source potential V_s, we can compute the surface potential at the source Φ_s for a given Q_s by inverting this integration. Once we know Φ_s, the depletion-layer width and depletion-layer charge can be calculated

from strictly electrostatic considerations, given the substrate doping level. From the surface potential and depletion-layer charge we can then determine the gate potential V_g. We obtain the channel current I by integrating the current-flow equations from one end of the channel to the other, using Q_s as a boundary condition. Thus, for each choice of mobile-charge density, we can separately compute the gate voltage and the corresponding channel current.

The more involved treatment of the transistor when it is not in saturation is an extension of the saturation case. Here, the mobile-charge density at the drain end of the channel Q_d is set not by velocity saturation, but by a boundary condition involving the drain voltage. Using this condition, we can build up a complete model for the transistor, covering all regimes of operation. The characteristics are completely continuous above and below threshold, in and out of saturation. This treatment takes into account all the effects of mobile-carrier velocity saturation.

Source Mobile-Charge Boundary Condition

The mobile charge per unit area in the channel region Q_m is a function of the distance z along the channel. At the barrier maximum just into the channel from the source, the boundary condition on the mobile-charge density ρ_s is given by the integral of the carrier density in the source region (a Fermi distribution) times the density of states $N(E)$ in the channel:

$$\rho_s = -q \int_{V_s - \Phi_s}^{\infty} \left(N(E) \frac{1}{e^E + 1} \right) dE.$$

For any realizable bias conditions, even for submicron devices, the source Fermi level is always many kT below the surface potential at the barrier maximum. The usual treatment, in which the Fermi function is replaced by a Boltzmann approximation, is thus valid. The resulting expression can be written as

$$\rho_s = -q N_{\text{eff}} e^{(\Phi_s - V_s)}.$$

The effective density of states in the channel N_{eff} is given by

$$N_{\text{eff}} = \int_0^{\infty} N(E) e^{-E} dE.$$

In these expressions, all energies are in units of kT, and hence all potentials are in units of kT/q.

The charge density Q_s is the integral from bulk to surface of ρ_s. The charge is actually quantum mechanically distributed, but we will assume that the charge is located at the surface. Then,

$$Q_s = -qN_{\text{eff}}e^{(\Phi_s - V_s)}. \tag{1}$$

Solving equation 1 for Φ_s and expressing Φ_s in ordinary volts:

$$\Phi_s = \frac{kT}{q}\ln\left(\frac{-Q_s}{qN_{\text{eff}}}\right) + V_s. \tag{2}$$

Note that the source voltage is referred to flat-band rather than to substrate Fermi level, so the junction band-bending must be added to the actual applied voltage.

Electrostatics

The complete electrostatics of the MOS device involves three independent potentials (source, drain, and gate) relative to substrate. We observe that the current through the channel always is controlled by the point along the channel where the potential barrier is maximum. This point is very near the source except when the voltage drop along the channel is nearly zero. Conditions on either side of this maximum point become progressively less important in determining the current. Because the potential is a maximum, one can accurately determine the solution normal to the channel using a one-dimensional analysis. The level of approximation used throughout this paper is to extend the conditions found from this one-dimensional solution toward the drain until the drain depletion layer is encountered. This approach factors an otherwise intractable problem into simple sub-problems that can be solved separately.

Figure 1 is a visualization of the potential distribution in the channel and shows the overall coordinate system. Figure 2 shows a cross-section through the barrier maximum of an MOS transistor (shown as n-channel). We assume that the substrate is uniformly doped with N acceptors per unit volume and hence the depletion layer contains a constant charge density $\rho = -qN$. In our coordinate system, x is measured perpendicular to the surface, with $x = 0$ at the substrate edge of the depletion layer. By simple application of Gauss' law, the

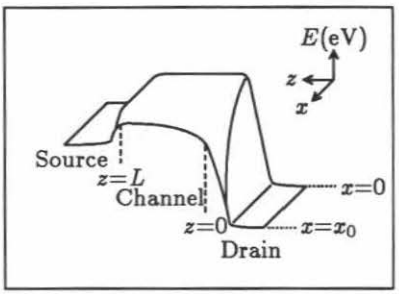

Figure 1: Potential distribution in the channel. (Adapted from Pao and Sah [8])

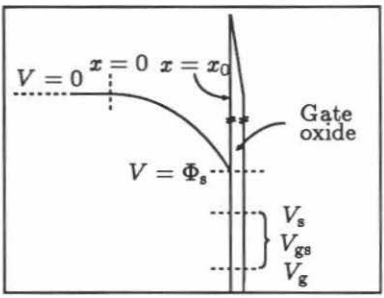

Figure 2: Cross-section normal to surface through the potential maximum ($z = l$).

electric field at any x is just equal to the total charge per unit area between the edge of the depletion layer and the point x, divided by ϵ_s, the permittivity of the semiconductor. The surface potential Φ is obtained by integrating this electric field from $x = 0$ to the surface $x = x_0$. The result of this integration is

$$\Phi = -\frac{1}{\epsilon_s}\frac{\rho x_0^2}{2}. \tag{3}$$

We will use the band edge deep in the substrate as the reference for all potentials. In this way, the surface potential is zero at flat-band. For any value of surface potential, equation 3 gives the depletion-layer thickness. Using this thickness, the total charge per unit area Q_{dep} uncovered in the depletion layer is

$$Q_{\text{dep}} = \rho x_0. \tag{4}$$

The voltage across the gate oxide is the electric field times the oxide thickness t_{ox}. The electric field is just the total charge divided by the permittivity of the oxide. The total charge is comprised of the depletion layer charge Q_{dep}, the mobile charge per unit area Q_{m}, and the surface fixed charge Q_{ss} (which includes the interface charge Q_{it} and any threshold adjustment charge). Consequently,

$$\begin{aligned} V_g &= \Phi - \frac{t_{\text{ox}}}{\epsilon_{\text{ox}}}(Q_{\text{tot}}) \\ &= \Phi - \frac{1}{C_{\text{ox}}}(Q_{\text{dep}} + Q_{\text{m}} + Q_{\text{ss}}). \end{aligned} \tag{5}$$

Given the surface potential, the gate voltage can be determined from equations 2, 3, 4 and 5. In general, the problem of deriving a self-consistent solution is difficult because the surface potential depends on mobile-charge density through equation 5, and the mobile-charge density depends on surface potential through the current flow equations. In addition, the boundary conditions on the mobile-charge density depend on surface potential through the Fermi distribution.

We can extract useful information from equation 5 for small changes in voltage around some operating point. Differentiating equation 5 with respect to Φ, we obtain

$$\frac{\partial V_g}{\partial \Phi} = 1 - \frac{C_{dep}}{C_{ox}} - \frac{1}{C_{ox}} \frac{\partial Q_m}{\partial \Phi}, \tag{6}$$

where

$$C_{dep} = \frac{\partial Q_{dep}}{\partial \Phi} = \frac{\epsilon_s}{x_0}, \qquad \text{and} \qquad C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}.$$

A particularly interesting simplification of the analysis can be derived from equation 6. For any given operating point, the gate is an equipotential and hence V_g does not depend on the coordinate z along the channel, whereas the surface potential Φ changes considerably. For the purpose of evaluating Φ, the lefthand side vanishes and we can define an *effective channel capacitance* C per unit area:

$$C = \frac{\partial Q_m}{\partial \Phi} = C_{dep} + C_{ox}. \tag{7}$$

Intuitively, the mobile charge is fixed by a boundary condition at the source. As it flows through the channel, there is a fixed relation between mobile charge and surface potential given by equation 7. In general, the capacitance C is a weak function of z, as is the mobility μ. We will first derive the zero-order result, taking C as constant and equal to the value at the potential maximum. This approximation is much less restrictive than the usual gradual channel approximation.

Channel Current

From the boundary conditions on mobile-charge density and surface potential at the source, we can now evaluate the channel current I. The current is dominated by diffusion under some circumstances, whereas in other regimes of operation the same transistor has charge carriers with drift velocities near saturation over the entire length of

the channel. We represent the current flow by a drift term and a diffusion term, and include the effects of velocity saturation in the drift term. Thus,

$$\frac{I}{w} = Q_{\mathrm{m}} v_{\mathrm{drift}} - \frac{kT}{q}\mu\frac{\partial Q_{\mathrm{m}}}{\partial z}, \tag{8}$$

where w is the width of the channel. The detailed functional form of drift velocity in the channel is not known with certainty. We adopt a simple relation that has the correct behavior at both high and low fields [4]:

$$v_{\mathrm{drift}} = v_0 \left(\frac{\mu E}{v_0 + \mu E}\right). \tag{9}$$

We now introduce a set of natural units, which we will use throughout the rest of this paper:

Velocity	v_0
Voltage	$\dfrac{kT}{q}$
Length	$l_0 = \dfrac{D}{v_0} = \dfrac{\mu kT}{v_0 q}$
Charge	$Q_{\mathrm{T}} = \dfrac{kT}{q}C$
Current	$v_0 Q_{\mathrm{T}}$

We define the *thermal charge* $Q_{\mathrm{T}} = CkT/q$ as the mobile charge per unit area at the potential maximum required to change the surface potential by exactly kT/q. The length unit l_0 can be thought of as a mean free path for electrons. All variables will be written in terms of these units, resulting in a dimensionless form for all equations.

In what follows, we will compute all currents for a channel of unit width. Because

$$\frac{\partial Q_{\mathrm{m}}}{\partial z} = \frac{\partial Q_{\mathrm{m}}}{\partial \Phi}\frac{\partial \Phi}{\partial z} = C\frac{\partial \Phi}{\partial z}$$

and

$$E = -\frac{\partial \Phi}{\partial z},$$

equation 8 can be written (in natural units) as

$$I = \frac{Q_{\mathrm{m}} Q'_{\mathrm{m}}}{Q'_{\mathrm{m}} + 1} + Q'_{\mathrm{m}},$$

or

$$I(Q'_{\mathrm{m}} + 1) = Q_{\mathrm{m}} Q'_{\mathrm{m}} + Q'_{\mathrm{m}}(Q'_{\mathrm{m}} + 1), \tag{10}$$

where the prime indicates derivative with respect to z, the distance along the channel. The first term on the righthand side of the equation is the drift term, and the second is the diffusion term. We will assume that the $(Q'_m)^2$ term is negligible compared to either the $Q_m Q'_m$ term (when Q_m is large) or the Q'_m term (when Q_m is small). This approximation is excellent as long as $l/l_0 \gg 1$. For a typical n-channel process, $l_0 \approx 0.015$ microns. Equation 10 can thus be written

$$\begin{aligned} I(Q'_m + 1) &= Q_m Q'_m + Q'_m \\ &= Q'_m(Q_m + 1) \end{aligned}$$

or

$$\begin{aligned} I &= Q_m Q'_m + Q'_m(1 - I) \\ &= \frac{\partial}{\partial z}\left(\frac{Q_m^2}{2} + Q_m(1 - I)\right). \end{aligned}$$

We now integrate both sides of this expression along the channel from drain ($z = 0$) to source ($z = l$). Noting that I is not a function of z,

$$Il = \frac{Q_s^2 - Q_d^2}{2} + (Q_s - Q_d)(1 - I). \tag{11}$$

A number of important insights into the operation of MOS devices can be gained from equation 11. For sufficiently large l, the current is small compared with unity, and $1 - I \approx 1$. This approximation corresponds to the usual treatment, ignoring velocity-saturation effects. Tracing through the derivation, we see that the quadratic term comes from the drift term in equation 8, and the linear term comes from the diffusion term in equation 8. The two terms make approximately equal contributions to the saturation current for $Q_s = Q_T$. For larger Q_s, the surface potential is dominated by mobile charge; for smaller Q_s, the surface potential is determined by the charge in the depletion layer. The condition $Q_s = Q_T$ corresponds to the common notion of *threshold*. We conclude that *below threshold, current flows by diffusion; above threshold, current flows by drift*, and at threshold, there is no discontinuity. The threshold shift due to the source substrate bias is modeled directly because the source potential in equation 2 need not be zero. No new terms need to be added to equation 5, and no new parameters need to be added to the model.

For a transistor in saturation, the charge density Q_d at the drain is moving at saturated velocity v_0. In natural units, this condition can be written $Q_d \approx I$. Equation 11 then becomes a simple quadratic in Q_s, the mobile charge density at the source, giving,

$$2Il + 1 = (Q_s + 1 - I)^2. \tag{12}$$

The solution to equation 12 is

$$\begin{aligned} I_{sat} &= Q_s + (l+1)\left(1 - \sqrt{1 + 2Q_s \frac{l}{(l+1)^2}}\right) \\ &\approx Q_s + (l+1)\left(1 - \sqrt{1 + \frac{2Q_s}{l}}\right). \end{aligned} \tag{13}$$

For sufficiently low drain voltages, the mobile charge at the drain Q_d is no longer moving at saturated velocity. Solving equation 11 explicitly for I gives

$$I = \frac{Q_s - Q_d}{Q_s - Q_d + l}\left(\frac{Q_s + Q_d}{2} + 1\right). \tag{14}$$

The effects of velocity saturation can be seen in equation 14. If $l \gg Q_s - Q_d$, then we can ignore the $Q_s - Q_d$ term in the denominator, and we have

$$I \propto \frac{Q_s^2}{2} + Q_s. \tag{15}$$

For large gate voltages, $Q_s \propto V_{gs} - V_{th}$ and the Q_s^2 term dominates. We have the familiar long-channel behavior:

$$I \propto (V_{gs} - V_{th})^2.$$

For gate voltages below threshold, the Q_s term dominates in equation 15, the charge is exponential in the gate voltage, and

$$I \propto e^{kV_{gs}}.$$

In the limit of velocity saturation, $l \ll Q_s - Q_d$, and we can ignore the l in the denominator of equation 14. Then the first fraction reduces to 1, and for large gate voltages

$$I \propto V_{gs} - V_{th}.$$

So, for a highly velocity-saturated device, there is a linear dependence of current on gate voltage.

Characteristics Below Saturation

We must now determine the boundary condition at the drain in order to evaluate Q_d as a function of Φ_d. We will define at every point along the channel a quasi-Fermi level or *imref* [9] ξ such that

$$\Phi - \xi = \frac{kT}{q} \ln\left(\frac{-Q}{qN_{\text{eff}}}\right). \tag{16}$$

This expression is, of course, just a generalization of equation 2. Writing equation 16 for both source and drain, assuming $\xi_s = V_s$ at the source, and subtracting the two expressions yields a relation between the surface potentials at the source and drain (Φ_s and Φ_d), the mobile-carrier densities at the source and drain (Q_s and Q_d), and the imrefs at the source and drain (V_s and ξ_d). We have

$$\Phi_d - \Phi_s + V_s - \xi_d = \frac{kT}{q} \ln\left(\frac{Q_d}{Q_s}\right). \tag{17}$$

We further assume, for the purpose of estimating the effect of small drain voltages on Q_d, that carriers at the drain end of the channel are Boltzmann distributed in energy with the same temperature as carriers in the drain. This approximation is exact in the limit of zero drain–source voltage. It will become less accurate when carriers are moving with saturated velocity.

We will derive the drain boundary condition by the following somewhat intuitive argument. Let the density of states in the drain be N_d and the density of states at the drain end of the channel be N_c. The probability P_{cd} of a carrier in the channel making a transition to a state in the drain is just the probability P_c of the state in the channel being occupied multiplied by the probability $1 - P_d$ that the corresponding state in the drain is unoccupied. A similar argument produces the probability that a carrier in the drain makes a transition back into the channel. Then,

$$P_{cd} = N_c P_c N_d (1 - P_d)$$

and

$$P_{dc} = N_d P_d N_c (1 - P_c).$$

The net drain current is proportional to the difference between these two probabilities:

$$P_{\text{net}} = P_{cd} - P_{dc} = N_c N_d (P_c - P_d). \tag{18}$$

Substituting P_c and P_d in terms of the *imrefs* as given in equation 16, equation 18 becomes

$$I = KQ_d \left(1 - e^{\xi_d - V_d}\right). \tag{19}$$

We notice that the constant K can be evaluated by considering operation at large drain voltages ($V_d \gg \xi_d$). This condition corresponds to saturation, with carriers at the drain end of the channel moving at saturated velocity. In natural units, this condition is written $Q_d = I$, and therefore $K = 1$. Consequently, equation 19 can be expressed as

$$\ln\left(1 - \frac{I}{Q_d}\right) = \xi_d - V_d. \tag{20}$$

Because equation 5 is valid for any surface potential, we can use equations 3, 4, and 5 to solve for Φ_d, yielding,

$$\Phi_d = \left(\sqrt{\frac{\epsilon_s \rho}{2C_{ox}^2} + V_g - \frac{Q_d}{C_{ox}}} - \sqrt{\frac{\epsilon_s \rho}{2C_{ox}^2}}\right)^2. \tag{21}$$

Substituting equation 20 into equation 17, we arrive at the final form of the relation between carrier density, current, and drain voltage:

$$V_d - V_s = \Phi_d - \Phi_s + \ln\frac{Q_s}{Q_d} - \ln\left(1 - \frac{I}{Q_d}\right). \tag{22}$$

The $\Phi_d - \Phi_s$ term is just the difference in the imrefs at the two ends of the channel. The $\ln(1 - I/Q_d)$ term is due to the "drain drop"; that is, the difference between ξ_d and V_d. The actual current for any given operating point can be found by simultaneous solution of equations 22 and 14.

Model Evaluation

In order to generate model curves for comparison with experimental data, the following algorithm was used. Voltages V_g and V_s were used with equations 2 and 5 to determine the source charge Q_s. Then the drain voltage V_d was used with equations 14, 21 and 22 to determine the drain charge and the drain current. Alternatively, several values of the drain charge were chosen varying between $Q_d = Q_s$ ($I_{ds} = 0, V_{ds} = 0$) and $Q_d = I_{sat}$ ($I = I_{sat}$, large V_{ds}), sweeping out the drain characteristic.

Experimental Results

We compared the model with a number of experimental devices with oxide thickness ≈ 100 Å and channel lengths down to $0.35\,\mu$, provided by Intel corporation. Detailed comparisons were made for devices from the same wafer, all of width $50\,\mu$ and of length ranging from $50\,\mu$ to $0.35\,\mu$ [7]. Mobility was taken from the channel conductance of the $50\,\mu$ device at very low drain–source voltage. Channel lengths were determined by comparing the channel conductance of a given device to that of the $50\,\mu$ device. Oxide thickness was obtained from the capacitance of a large MOS-dot. Substrate doping was found by plotting the threshold voltage versus the square root of the source–substrate reverse bias, as shown in Figure 3. The fixed charge at the surface Q_{ss} was computed directly from the threshold voltage once the substrate doping was known. This charge includes any threshold-adjustment implant dose. The saturated velocity of electrons in silicon was taken from the literature [6].

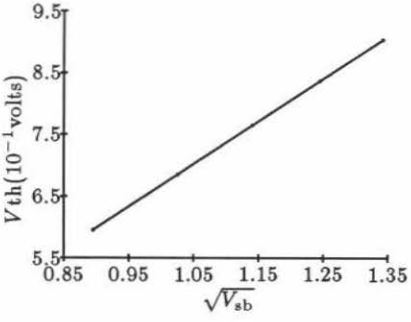

Figure 3: Determination of substrate doping $N_d = 10^{17}/cm^3$.

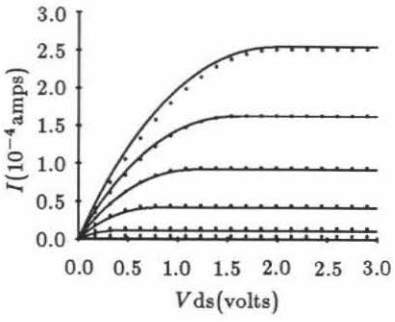

Figure 4: Drain current vs. V_{ds} for fixed V_{gs} for a $50\,\mu$ transistor.

The results of comparing the zero-order model for $50\,\mu$, $0.7\,\mu$, and $0.35\,\mu$ devices are shown in Figures 4, 5, and 6. Drain curves are shown for fixed V_{gs} ranging from 0 to 3 volts. The theoretical curves use the same set of parameters in all cases, except that the value of Q_{ss} was found to be slightly larger (in the direction to increase the threshold) for shorter devices. Any minute lateral surface diffusion during drain–source drive could easily produce such an effect. Although the threshold shift induced by this effect was small, a correction was made for each length in order to fit the subthreshold current, which is exponential in Q_{ss}. Note that this effect is in the opposite direction from

the commonly expressed notion that threshold voltages decrease with decreasing length. In any case, the agreement is quite good. The model is simple to evaluate and the magnitudes of the curves match well. The results are certainly adequate for most digital applications.

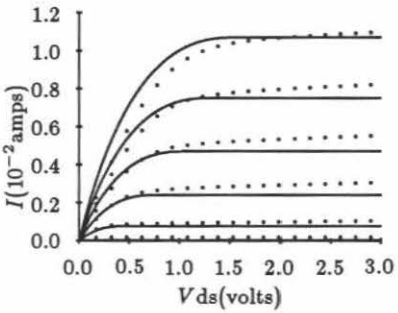

Figure 5: Drain current vs. V_{ds} for fixed V_{gs} for a 0.7 μ transistor.

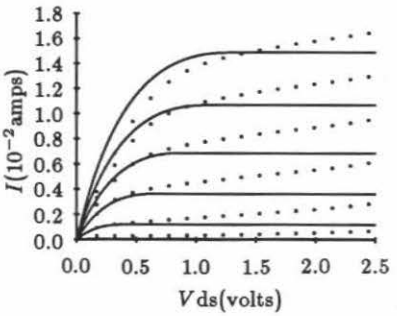

Figure 6: Drain current vs. V_{ds} for fixed V_{gs} for a 0.35 μ transistor.

First-Order Corrections to the Basic Model

There are several first-order effects, the consequences of which can be seen in Figure 6. The slope of the drain curves in saturation has not yet been considered. This dependence of saturation current on drain voltage is due to the change in channel length l with drain bias in equation 14. This well-known behavior is called the Early effect, after Jim Early who first explained the phenomenon in bipolar transistors [2].

The conductance of the actual device near the origin is less than that predicted by the model, and the discrepancy is larger for larger gate voltages. This behavior is due to the dependence of mobility on electric field perpendicular to the direction of current flow. Intuitively, the field from the gate attracts electrons in the channel toward the oxide interface. Conditions at this interface are not as ideal as they are in the silicon crystal, and an electron is more likely to be scattered if it spends more time there. This additional scattering decreases the electron's mean free time, and hence reduces its mobility.

Another effect is that the experimental saturation currents are less dependent on gate voltage than would be expected. This discrepancy

is due to parasitic resistances of the source and drain. Although resistance is not a device property in the strictest sense, it is a necessary and unavoidable byproduct of any real fabrication process. As channel lengths are made shorter, control can be maintained only by reducing the depth of source and drain diffusions. Shorter channels have less resistance of their own, but are necessarily accompanied by larger and larger sheet resistance in source and drain. The ratio of the resistances thus scales as the square of the channel length.

The several effects mentioned, along with the first-order corrections to the model itself, are of roughly the same magnitude. Some, such as the Early effect, increase the current. Others, such as mobility variation and internal resistances, decrease the current. Our *modis operandi* is to find a particular regime of operation in which one of the effects is dominant, and to evaluate the effect there. We study each effect where it can be isolated and analyzed independently.

Mobility Variation

The vertical electric field acting perpendicular to the channel leads to mobility degradation with increasing V_{gs}. This effect can be seen quite clearly in a 50μ transistor in which velocity saturation and series resistance are negligible. A plot of the low-field channel conductance versus gate voltage is shown in Figure 7. Also shown in Figure 7 is the derivative of the conductance curve $\frac{\partial G}{\partial V_{gs}}$. If the mobility were constant, the conductance plot would be a straight line with x-intercept at threshold, and the derivative would be constant above threshold. The slope of the derivative curve is a direct measurement of the mobility variation with the gate electric field, E_g. By adding another term to the scattering model used to derive the velocity saturation, we obtain a form for μ:

$$\mu_{eff} = \frac{\mu}{1 + \frac{E_g}{E_f}} \tag{23}$$

Mobility variation was added to the model by replacing μ in equation 9 with μ_{eff}. The vertical electric field E_g was calculated from the relation

$$E_g = \frac{Q_{tot}}{\epsilon_s} \tag{24}$$

where Q_{tot} is the total charge as used in equation 5. The parameters μ and E_f can be evaluated directly from the data in Figure 7. The value of the mobility was found to be 490 $\text{cm}^2/\text{volt} - \text{sec}$. The results of

this refinement to the model and of the uncorrected model are plotted along with the original experimental data.

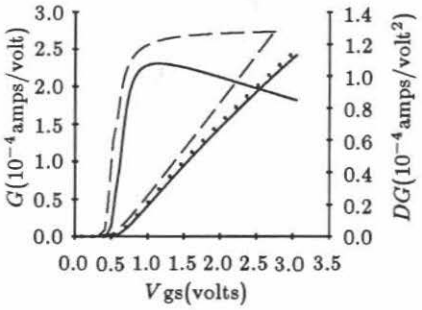

Figure 7: Conductance and slope of conductance for a $50\,\mu$ device. Dots: data; solid lines: corrected model curves; dotted lines: model without mobility variation.

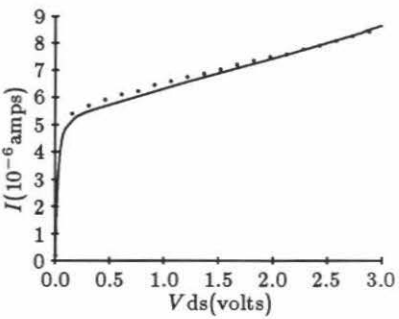

Figure 8: Drain current vs. V_{ds} for $V_{gs} = 0.5V$ for a $0.7\,\mu$ transistor. Dots: data; solid line: corrected model curve.

Early Effect

The Early effect (drain-voltage modulation of channel length) is important in today's devices, and becomes crucial as devices scale to submicron lengths. The effect is best observed in a short-channel device in subthreshold, where there are no mobile charge carriers to reduce the effect. Current flows by pure diffusion so there is no velocity saturation in the channel proper. We know the surface potential at the source edge of the channel, and know that it is constant to the very edge of the drain depletion region. Hence, a measurement of the slope of saturation current amounts to a direct measurement of the change in channel length with drain voltage:

$$\frac{\partial I_{sat}}{\partial V_d} = \frac{1}{l}\frac{\partial l}{\partial V_d} I_0,$$

where I_0 is the saturation current in the absence of the Early effect. An example of the direct manifestation of the Early effect can be seen in the subthreshold current of a 0.7 μ device in Figure 8.

To incorporate the Early effect into our model, the effective channel length is calculated by subtracting the lengths of the depletion layers

at source and drain from the physical length:

$$L_{\text{eff}} = L_{\text{phys}} - \delta L_{\text{source}} - \delta L_{\text{drain}}.$$

The actual boundary around the depletion layer near the drain is a complicated, two-dimensional affair, involving not only the the fixed charges in the depletion layer but also the mobile charge in the channel [1,3]. We used the source and drain surface potentials, electric fields, and voltages as the boundary conditions in the solution of Gauss' Law. A simple cylindrical approximation to the two-dimensional solution around the drain "corner" in subthreshold gives a value for δL that is $\sqrt{2}$ times as large as that predicted for a planar junction. The factor derived from Figure 8 is 1.4. To add the Early effect to our model the value of l in equation 14 was replaced by L_{eff}. The result of this correction is the model curve shown in Figure 8.

Electric-field lines from the drain can terminate on mobile electrons as well as on the negative fixed charges in the depletion layer The density of mobile charge increases at higher gate voltages. We therefore expect a smaller change in channel length at high gate voltages than in subthreshold. A graphic illustration of the effect of mobile charge on the Early effect can be seen in Figure 9, which is a drain curve for higher gate voltage (3.0 V, well above threshold). The theoretical curve shown is that predicted by the Early effect, ignoring the contribution of mobile charge. The corrected model curve is shown as part of the drain characteristics in Figure 13.

We calculate an approximate volume charge density by normalizing the mobile charge/unit area by the width of depletion layer normal to the surface. So ρ now becomes

$$\rho_{\text{eff}} = \rho_{\text{mobile}} + \rho_{\text{depletion}}.$$

In the model calculation, the vertical depletion layer widths at the source and drain ends of the channel were used as normalizing factors. These values were calculated from equation 4. The value of ρ_{eff} was used in place of ρ in the calculation of the depletion-layer length.

Since the mobile charge increases the total charge in the electrostatic equations, it decreases the extent to which the saturation current changes with drain voltage. This decrease in Early effect with mobile-charge density we have called the *Late effect.* It is clear that the Late effect makes the device a better current source and hence, in

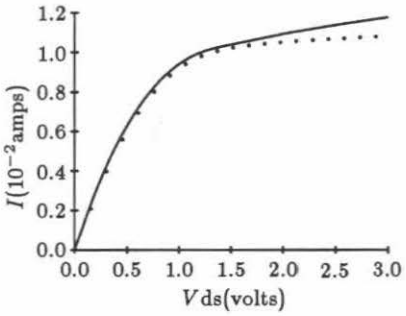

Figure 9: Drain Current vs. V_{ds} for $V_{gs} = 3V$ for a $0.7\,\mu$ transistor. Dots: data; solid line: model curve without mobile charge effect.

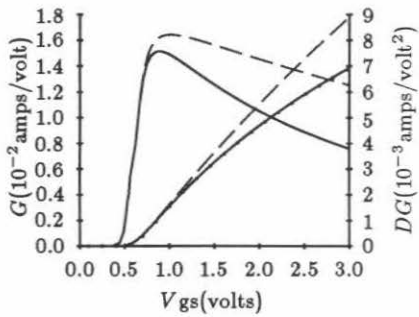

Figure 10: Conductance and slope of conductance for a $0.7\,\mu$ device. Dots: data; solid lines: corrected model curves, dashed lines: model without resistance.

some sense, more ideal. However, the effect has decreased the current-driving capability of the device.

Resistance

Although not strictly part of a device model, source and drain resistance must be included to compare the model to any real measurements. Direct measurements of the sheet resistance of the diffusion layer gave 68 ohms per square. In the test devices, the distance from the metal contact cuts to the edge of the gate was 5 μ. The devices were 50 μ wide, so the source and drain resistances were 7 Ω. For a short, wide device 50 μ by 7 μ we measured drain currents up to 10 mA. The corresponding voltage drop across the source resistor amounted to an error of 70 mV, about 15 per cent of the spacing between adjacent drain curves. To compute accurate device characteristics, source and drain resistors were added to the model. This equivalent circuit was simulated by iteration, alternately updating V_s and V_d and evaluating the model. Figure 10 shows the experimental data and the model results with and without resistance.

The Upgraded Model

In addition to the phenomena described above, two other first-order effects were added to the model. The width of the depletion layer in the channel increases with distance towards the drain. This effect corresponds to a decrease in the capacitance C in equation 7. The vertical electric field E_g also varies along the channel, reaching a minimum at the drain, affecting the mobility variation in equation 23 at high drain voltages. Both of these variations were interpolated linearly between the known values at source and drain. The effects mentioned were incorporated into a unified model, which was used to generate theoretical curves for all measured properties of transistors of a wide range of lengths. The family of measurements for 50 μ, 2 μ, 0.7 μ, and 0.35 μ devices are shown in Figures 11, 12, 13, and 14. For each device, we show the drain characteristics for gate voltages of 0 to 3 Volts in 0.5 Volt steps.

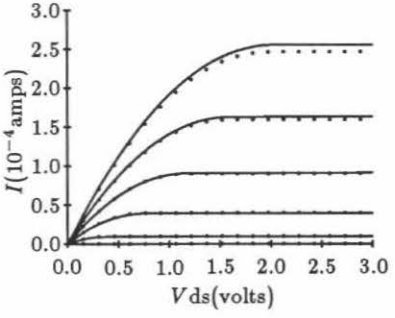

Figure 11: Drain Current vs. V_{ds} for $V_{gs} = 0 - 3\ V$ for a $50\,\mu$ transistor.

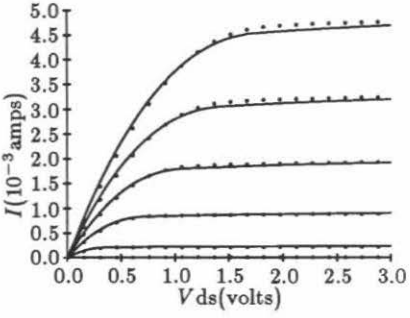

Figure 12: Drain Current vs. V_{ds} for $V_{gs} = 0 - 3\ V$ for a $2\,\mu$ transistor.

Conclusions

It can be seen that the model generates curves that are in excellent agreement with experiment over a wide range of device sizes without resorting to *ad hoc* parameters. All parameters either are derived from the process by direct measurement, or are physical dimensions of the layout of a particular device. We believe these results demonstrate that a simple first-principles model with physically meaningful, measurable parameters is quite capable of quantitatively predicting the behavior of MOS devices down to their limit of usefulness [5].

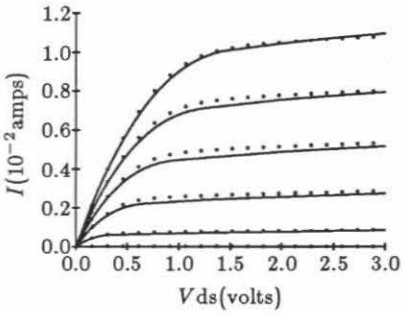

Figure 13: Drain Current vs. V_{ds} for $V_{gs} = 0 - 3\ V$ for a $0.7\,\mu$ transistor.

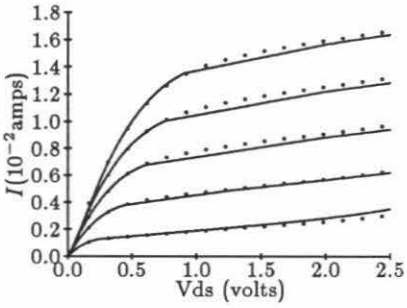

Figure 14: Drain Current vs. V_{ds} for $V_{gs} = 0 - 3\ V$ for a $0.35\,\mu$ transistor.

Acknowledgments

This work was supported by the System Development Foundation and the Semiconductor Research Corporation. Experimental devices were kindly provided by Amr Mohsen and Gerry Parker of Intel Corporation. We are grateful to Massimo Sivilotti, David Feinstein, and Cecilia Shen for many valuable discussions. We also thank Jim Campbell and Kathy Doughty for their help in acquiring experimental data, and Cal Jackson, Michael Emerling, Dave Gillespie, and Glenn Gribble for their aid in preparing this manuscript.

References

[1] G. Baum and H. Beneking. Drift velocity saturation in MOS transistors. *Correspondance IEEE Transactions on Electron Devices*, ED-17, 1970.

[2] J. M. Early. Effects of space charge layer widening in junction transistors. *Proceedings of the IRE*, 40, 1952.

[3] D. Frohman-Bentchkowsky and A. S. Grove. Conductance of MOS transistors in saturation. *IEEE Transactions on Electron Devices*, ED-16, 1969.

[4] B. Hoeneisen and C. A. Mead. Current-voltage characteristics of small size MOS transistors. *IEEE Transactions on Electron Devices*, ED-19, 1972.

[5] B. Hoeneisen and C. A. Mead. Fundamental limitations in microelectronics — I. MOS technology. *Solid State Electronics*, 15, 1972.

[6] C. C. Jacoboni, G. Canali, and A. A. Quaranta. A review of some charge transport properties of silicon. *Solid State Electronics*, 20, 1977.

[7] C. A. Mead and M. C. Maher. A charge controlled model for submicron MOS. In *Proceedings of the Colorado Microelectronics Conference CMC-86*, 1986.

[8] H. C. Pao and C. T. Sah. Effects of diffusion current on the characteristics of metal-oxide (insulator)-semiconductor transistors. *Solid State Electronics*, 9, 1966.

[9] W. Shockley. *Electrons and Holes in Semiconductors.* D. Van Nostrand, 1960.

[10] Y. Tsividis and G. Masetti. Problems in precision modeling of the MOS transistor for analog applications. *IEEE Transactions on Computer-Aided Design*, CAD-3, 1984. An excellent recent review of MOS models, together with a comprehensive set of references.